

映像の種類に依存しない映像要約手法に関する検討

倪 嬋斌[†] 渡部秀一[‡] 野村敏男[‡] 亀山渉[‡][†]早稲田大学大学院国際情報通信研究科[‡]シャープ株式会社

1. はじめに

ここ数年、私たちの周りでは利用できるマルチメディア情報が飛躍的に増大している。しかし、これらの膨大なマルチメディア情報からユーザが欲する情報を汲み取るための環境は未だ整っていない現状である。近年では EPG（電子番組表）に含まれるキーワードを基に自動録画する機能を用いて番組単位の検索は可能となったが、録画したコンテンツの中から、興味のある箇所だけを見たいといったニーズへの対応は果たされていない。そこで本稿では、ユーザの所望する映像区間を符号化情報パラメータにおける特徴パターンに基づいて判定する方式を検討した。

2. 従来手法の問題点と提案手法

マルチメディアコンテンツに関する内容情報（メタデータ）の記述方式として、MPEG-7 や TV-Anytime などが策定されているが、多種多様な映像データの細かい事象への対応や、すでに流通している映像データにメタデータを付与する作業は高コストである。また、これらの規格では様々な記述ツールが用意されており、その多くが人手によって与えるハイレベルなテキスト情報によるものであり、一般ユーザがコンテンツ内容を明確な語句に表現できないという問題点がある。そのため、コンテンツのストリーム信号からローレベルなデータ群を抽出、解析することで、テキスト情報では表現できないような情報を定義・表現することは、メディア理解・認識の分野における継続的な研究課題である[1]。

いままで符号化情報を利用して映像の特徴を認識・理解しようとする方法はすでに様々な方法が提案されている[2][3]。また、メディア理解による映像メディア構造化についての有用性が数多く報告されている。しかし、これらの研究成果とユーザ主観評価との対応関係についてはまだ研究の余地が残されている。

そこで本稿では、代表的な映像圧縮符号化方式である MPEG で符号化された映像情報中の情

報に注目し、各種符号化パラメータ情報を統計解析した結果と、ユーザ主観評価の対応関係に基づいた映像重要度を判定する方式について検討した。

また、映像に依存した事前の知識を用いて特定の種類の映像を要約する手法についてはいくつかの提案方式が存在しているが、これらの方式では、要約映像を作る対象が限定されるという問題がある。本稿では映像の種類に依存しない方法として、「映像の文法」[4]が存在しているという仮定の下、特定の知識を必要とすることなく部分部分の映像重要度を判定する方式を検討した。本稿では主として、この重要度に基づいて要約映像を生成する手法について、主観評価実験の結果とその考察について述べる。

3. 主観評価実験

シーンを最小単位とする映像区間の重要度を得るため、被験者を使った映像の主観評価実験を実施した。

3-1. 実験方法

被験者：男女大学院生 10 名（24 歳～28 歳）及び社会人 19 名（24 歳～43 歳）。

刺激：以下の 2 通りの刺激を用意した。刺激(1)：オリジナル映像を提示するもの。刺激(2)：予めシーン単位にカットした映像を順次提示するもの。なお、刺激(2)ではシーンを視聴した直後に質問票に答える時間(5 秒)を与える。また、被験者を刺激の映像的特徴に集中させるため、刺激の提示条件を無音条件で統制した。実験の試料としては、本稿のねらいである映像の種類に依存しないことを考量し、(1) 情報・紀行・ドキュメンタリ番組、(2) ニュース番組、(3) バラエティ番組、(4) スポーツ中継、(5) ドラマの 5 つの代表的なジャンルのテレビ番組映像を用いた。

計測：当該シーンの全体の理解においての影響度を「かなり影響する 3」、「影響する 2」、「やや影響する 1」、「どちらとも言えない 0」、「やや影響しない-1」、「影響しない-2」、「全く影響しない-3」の七件法で質問票を作成し、被験者に自己採点を行わせることとした。

実施手順：まず、被験者に対して、実験の全体の流れを説明し、刺激の提示の際には、映像全体のバランスを考えながら個々のカットの採点をするように教示を与えた。次に、各々の刺激を提示しながら質問票に自己採点を行わせるこ

とした。ジャンル毎に最低 10 分以上の休息を与えることとした。

3-2. 実験結果

図 1～図 5 に、実験結果を示す。なお、各図において、横軸、縦軸はそれぞれシーン番号、主観評価値の平均値及び誤差範囲を表す。

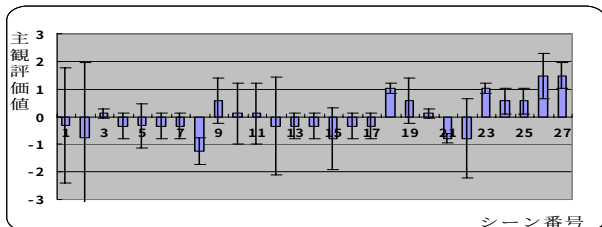


図 1. 情報・紀行・ドキュメンタリ番組

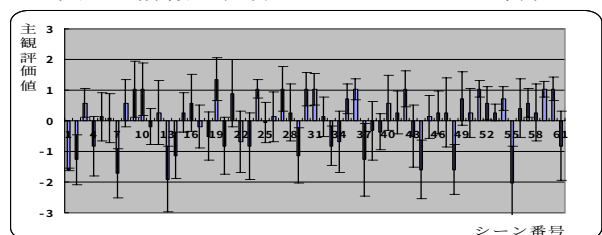


図 2. ニュース番組

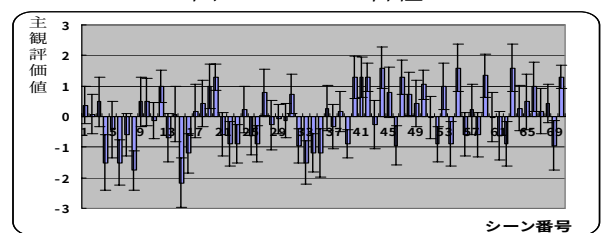


図 3. バラエティ番組

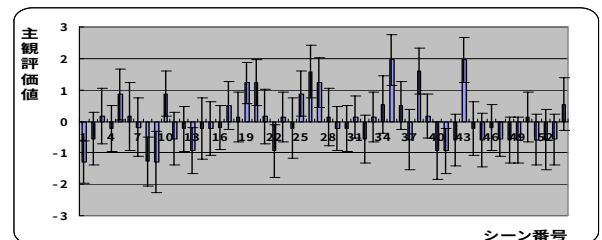


図 4. スポーツ中継

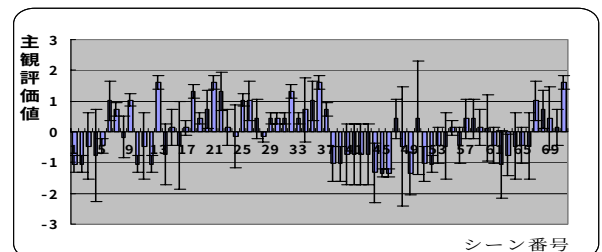


図 5. ドラマ

4. 考察

番組の映像的な特徴とは直接関係のない番組の内容や登場人物への嗜好などの普遍的でない情報が含まれるものの、全種類の映像コンテンツに高采点、低偏差のカットが存在することが

確認できた。また被験者の間の相関が強いことが分かった。採点結果や実験直後の内省報告により、全種類の映像コンテンツに共通して以下の要因が高採点の傾向に影響を及ぼすと思われる。

- (1) 制作者の意図が感じ取れるカメラワーク (PAN,ZOOM, FIX)。
- (2) 画面中に支配的なオブジェクトが存在する。
- (3) 人物が映っている。
- (4) 動きが激しい。
- (5) テロップがある。

この事実より、今回の実験で映像コンテンツユーザに共通する嗜好をある程度把握できたと思われる。

本研究の目的は、過度に重い処理をせずに符号化ビットストリームから得られる情報のみを元にシーンの重要度を判定する事であるので、MPEG のビットストリームに埋め込まれた種々の情報から、上記のような特徴量をどのように抽出するのかが今後の課題である。具体的には、主観評価値と符号化情報の統計処理によって、その対応関係を見出す事が可能であると考えられる。

5. まとめと今後の課題

本稿では、映像の種類に依存しない、映像中のシーン重要度を抽出する方法について検討した。主観評価の結果、シーンの重要度には個人差と映像の種類によらない明らかな傾向があった。その原因として、何かの特徴的な映像撮影と編集手法があることが伺え、いわゆる「映像の文法」がそれではないかと考えられることが分かった。今後は、主観評価と符号化情報の対応関係を明らかにし、「映像の文法」に基づいた映像要約手法の実現と、その有用性を明らかにする予定である。

参考文献

- [1]馬場口登, “メディア理解による映像メディアの構造化,” 信学技報, IE-9918 PRMU99-42 MVE99-38(1999-07),pp39-46,Jul.,1999.
- [2]杉山善明, 渋谷亜紀, 有木康雄, “DCT 特徴に基づく TV スポーツ映像の自動判別,” 信学技報, IE98-28, PRMU97-59, MVE97-44, pp33-38, 1997-07
- [3]青木真吾, 森田啓義, 荒俣吉壮, 西新幹彦 “マクロブロックタイプを用いた MPEG2 圧縮動画画像のカット点検出,” 情報処理学会論文誌 Vol.46 NO.SIG 15(CVIM 12), pp51-58
- [4]Daniel Arijonf, “Grammar of the Film Language,” London: Focal Press, 1976 (邦訳: 映画の文法)